

ENGG4920CP Thesis II

# ON MULTI-RATE SEQUENTIAL DATA TRANSMISSION

BY

Cheuk Ting LI

A FINAL YEAR PROJECT REPORT SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE  
OF BACHELOR OF INFORMATION ENGINEERING  
DEPARTMENT OF INFORMATION ENGINEERING THE  
CHINESE UNIVERSITY OF HONG KONG

May, 2012

# ON MULTI-RATE SEQUENTIAL DATA TRANSMISSION

CHEUK TING LI

**ABSTRACT.** In this report, we investigate the data transmission model in which a sequence of data is broadcasted to a number of receivers. The receivers, which have different channel capacities, wish to decode the data sequentially at different rates. Our results are applicable to a wide range of scenarios. For instance, it can be employed in the broadcast streaming of a video clip through the internet, so that receivers with different bandwidths can play the video at different speed. Receivers with greater bandwidths can provide a smooth playback, while receivers with smaller bandwidths can play the video at a slower speed, or with short pauses or rebuffering.

## 1. INTRODUCTION

Consider the scenario in which a long video clip has to be transmitted to a number of receivers having different packet loss ratios. One approach is to divide the video data into blocks of  $K$  packets, encode each block into  $L \geq K$  encoding packets, and then transmit the blocks to the receiver sequentially. Using random linear projections or any capacity-achieving erasure code, the receiver can decode the block if about  $K$  out of  $L$  packets are received. This method, which we call a blockwise code, can only cater for the need of the receiver with packet loss probability less than  $1 - K/L$ .

To suit the need of different receivers, we can perform time multiplexing on two blockwise codes at different rates. Consider Blockwise code 1 and Blockwise code 2, which use random linear projections to encode each block of  $K$  packets into  $L_1$  and  $L_2$  packets respectively ( $L_1 < L_2$ ). Denote the  $i$ -th packet generated using Blockwise code  $k$  by  $P_{k,i}$ . We transmit the packets of the two codes in an interleaved manner (in the sequence  $P_{1,1}, P_{2,1}, P_{1,2}, P_{2,2}, P_{1,3}, \dots$ ). Receiver 1, which uses only the packets generated using Blockwise code 1, can decode a block using  $K$  out of the  $L_1$  packets encoded from the block, and therefore can tolerate a packet loss probability  $1 - K/L_1$ . As Blockwise code 1 transmits a block of  $K$  packet per  $L_1$  channel uses, taking interleaving into account, Receiver 1 can decode at a rate of  $K/(2L_1)$  packets per channel use. Receiver 2 uses packets generated by both codes. It can decode a block using  $K$  out of the  $L_1 + L_2$  packets encoded from the block, and allows a higher packet loss probability  $1 - K/(L_1 + L_2)$ . However, to use the packets generated by both codes, Receiver 2 has to wait for the slower Blockwise code 2, which transmits a block of  $K$  packet per  $L_2$  channel uses. Receiver 2 can decode at a rate  $K/(2L_2)$ .

In the scenario, the receivers with different channel conditions wish to decode the same sequence of data. Each receiver will decode the data sequentially at a roughly constant rate which depends on the channel condition. We call these settings as *multi-rate sequential data transmission*. It can be viewed as multilevel diversity coding [1] with an additional sequential decoding constraint. In the following sections, we will discuss various cases of the problem. The case of one transmitter is described in Section 3. The main contribution lies in Section 6 concerning multiple transmitters, in which multi-rate sequential data transmission has apparently dissimilar behavior compared to its non-sequential counterpart.

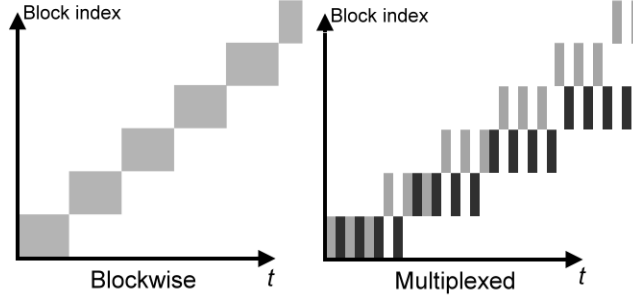


FIGURE 1.1. Encoding process of blockwise and multiplexed codes

## 2. FORMULATION OF MULTI-RATE SEQUENTIAL DATA TRANSMISSION

We consider the transmission of a sequence of data through an erasure channel. To simplify the setting, we assume that one bit is sent at a time through the erasure channel. Assume the sender is going to transmit a sequence of bits  $M_1, M_2, \dots$ . We assume  $M_i \stackrel{i.i.d.}{\sim} \text{Bern}(1/2)$ . The sender will encode it into a sequence of bits  $X_1, X_2, \dots$  and transmit them through an erasure channel. The symbols arrive at the receiver as  $Y_1, Y_2, \dots$ , where some of them may be erased (denoted by  $e$ ,  $Y_i \in \{0, 1, e\}$ ). Based on these symbols, the receiver tries to decode  $M_1, M_2, \dots$ .

**Definition 1** (MRS code). A *multi-rate sequential code* (MRS code) is specified by a pair of encoding and decoding functions. The encoding function is a function mapping the message  $\{M_i\}$  to the encoding symbols  $\{X_j\}$ . There is a random variable  $Q$  supported in  $\mathcal{Q}$  which is known by both the sender and the receiver, and independent of the message and the channel erasure, to allow random coding scheme. The encoding function is a function

$$\begin{aligned} \text{Enc} : \mathcal{Q} \times \{0, 1\}^{\mathbb{N}} \times \mathbb{N} &\rightarrow \{0, 1\} \\ (Q, \{M_i\}, j) &\mapsto X_j. \end{aligned}$$

Note that this definition allows the encoder to look at all blocks.

The decoding function maps the received symbols  $\{Y_j\}$ , where some of them may be erased, to the recovered blocks  $\{\widetilde{M}_i\}$ . Let  $\mathcal{Y} = \bigcup_{n \in \mathbb{N}} \{0, 1, e\}^n$  be the space of received symbols. The decoding function is a function

$$\begin{aligned} \text{Dec} : \mathcal{Q} \times \mathcal{Y} \times \mathbb{N} &\rightarrow \{0, 1\} \\ (Q, Y_1^n, i) &\mapsto \widetilde{M}_i. \end{aligned}$$

We use the notation  $Y_a^b = (Y_a, Y_{a+1}, \dots, Y_b)$ . For simplicity, we write  $X_j(Q, \{M_i\}) = \text{Enc}(Q, \{M_i\}, j)$ , and  $\widetilde{M}_i(Q, Y_1^n) = \text{Dec}(Q, Y_1^n, i)$ .

The MRS code does not admit a fixed rate like other block codes. Instead its rate depends on the channel capacity.

**Definition 2** (admissible pair). A rate-capacity pair  $(r, c)$  is called  $\epsilon$ -admissible by a code if there exist  $N_0$  such that when  $X_i \rightarrow Y_i$  is an erasure channel with capacity  $c$  (i.e. erasure channel with erasure probability  $1 - c$ ),

$$\mathbb{P} \left\{ M_m \neq \widetilde{M}_m(Q, Y_1^N) \right\} \leq \epsilon \text{ for any } N \geq N_0 \text{ and } m \leq N(r - \epsilon).$$

In other words, any receiver with channel capacity  $c$  can decode the first  $N(r - \epsilon)$  bits  $M_1^{N(r - \epsilon)}$  with bit error probability less than  $\epsilon$  when the first  $N$  symbols  $Y_1^N$  are received, for sufficiently large  $N$ .

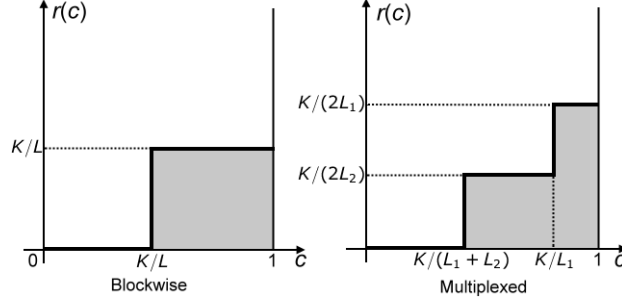


FIGURE 2.1. Rate-capacity functions of blockwise and multiplexed blockwise code

It is clear that if  $(c, r)$  is  $\epsilon$ -admissible, then all pairs in  $\{(c', r') | c' \geq c, r' \leq r\}$  are  $\epsilon$ -admissible. Therefore we can use a function to characterize the rate of a code. We call  $r : [0, 1] \rightarrow [0, \infty)$  a *rate-capacity function* if it is monotonically increasing, right continuous, and there exists an  $\eta > 0$  such that  $r(c) = 0$  for  $c \leq \eta$ .

**Definition 3** (rate of MRS code). A rate-capacity function  $r(c)$  is called  $\epsilon$ -admissible by a code if all of the pairs  $(c, r(c))$  are  $\epsilon$ -admissible by the code.

The rate-capacity functions of the blockwise code and the multiplexed blockwise code described in the introduction can be given by Figure 2.1.

**Definition 4** (achievable rate-capacity functions). A rate-capacity function  $r(c)$  is *achievable* if for any  $\epsilon > 0$ , there exist a code where  $r(c)$  is  $\epsilon$ -admissible by that code.

### 3. SUPERPOSITION CODING FOR SINGLE TRANSMITTER

In this section, we will present the design of superposition multi-rate sequential codes, and prove their optimality by giving the set of achievable rate-capacity functions explicitly.

**Definition 5** (superposition MRS code). A superposition MRS code is characterized by the block size  $K$  and the parameter  $g : [0, \infty) \rightarrow [0, \infty)$  which is bounded, monotonically decreasing and left continuous with  $\int_0^\infty g(\alpha) d\alpha = 1$ . The message  $\{M_i\}$  is divided into blocks of  $K$  bits,  $B_i = M_{(i-1)K+1}^K$ . In the encoding process of the code, we first sample a sequence of non-negative random variables  $A_1, A_2, \dots$  i.i.d. according to the cumulative distribution function

$$\begin{aligned}
 F_A(\alpha) &= \int_0^\alpha g(x) dx - \alpha g(\alpha) \\
 (3.1) \quad &= - \int_0^\alpha x dg(x).
 \end{aligned}$$

Note that  $F_A(\alpha)$  is increasing as  $g(\alpha)$  is decreasing. The sequence is known by both the sender and the receiver (we may let  $Q = (A_1, A_2, \dots)$ ). At time instance  $i$ , the sender generate an encoding symbol  $X_i$  from the block  $B_{\lceil i \cdot A_i / K \rceil}$  using random linear projections and transmit it to the receiver.

Note that we employ a random coding scheme. Random linear projections allow us to decode a block of  $K$  bits using  $K + o(K)$  encoding bits with an arbitrarily small error probability. The superposition MRS code is essentially performing time multiplex on blockwise codes at different rates.

The parameter  $g(\alpha)$  roughly corresponds to the proportion of the encoding symbols dedicated to satisfying the need of the receivers which wish to decode at rate  $\alpha$ . When we send encoding symbols encoded at rate  $\gamma$  (i.e.  $A_i = \gamma$ ), a receiver which decodes at rate  $\alpha \leq \gamma$  can use a portion of  $\alpha/\gamma$  of the encoding symbols. The parameter  $g(\alpha)$  describes the proportion of the symbols which can be used, divided by the proportion needed (which is  $\alpha$ ), given by

$$g(\alpha) = \frac{1}{\alpha} \int_{\alpha}^{\infty} \frac{\alpha}{\gamma} dF_A(\gamma),$$

which can be verified using (3.1):

$$\begin{aligned} \frac{1}{\alpha} \int_{\alpha}^{\infty} \frac{\alpha}{\gamma} dF_A(\gamma) &= \int_{\alpha}^{\infty} \frac{1}{\gamma} dF_A(\gamma) \\ &= - \int_{\alpha}^{\infty} \frac{1}{\gamma} \cdot \gamma dg(\gamma) \\ &= - \int_{\alpha}^{\infty} dg(\gamma) = g(\alpha). \end{aligned}$$

The blockwise code and the multiplexed blockwise code described in the introduction are examples of superposition MRS codes. For the blockwise code, the parameter is taken to be

$$g(\alpha) = \begin{cases} L/K & \text{when } \alpha \leq K/L \\ 0 & \text{when } \alpha > K/L. \end{cases}$$

For the multiplexed blockwise code, the parameter is taken to be

$$g(\alpha) = \begin{cases} \frac{L_1+L_2}{K} & \text{when } \alpha \leq \frac{K}{2L_2} \\ \frac{L_1}{K} & \text{when } \frac{K}{2L_2} < \alpha \leq \frac{K}{2L_1} \\ 0 & \text{when } \alpha > \frac{K}{2L_1}. \end{cases}$$

The choice of the parameter is closely related to the rate-capacity function we would like to achieve. The following theorem describes the relationship between the two functions.

**Theorem 6.** *For a fixed  $\epsilon$ , the rate-capacity function  $r(c)$  is  $\epsilon$ -admissible by the superposition MRS code with block size  $K$  and parameter  $g(\alpha)$  for all sufficiently large  $K$ , if there exists  $\xi > 0$  satisfying*

$$c \cdot g(r(c)) \geq 1 + \xi \text{ for all } c > 0 \text{ with } r(c) > 0.$$

*Proof.* Fix  $\epsilon > 0$ . Let  $r(c)$  be a rate-capacity function, and let  $g(\alpha)$  be a bounded and monotonically decreasing function (let  $g(\alpha) \leq \zeta$  for all  $\alpha \geq 0$ ). Assume the condition  $c \cdot g(r(c)) \geq 1 + \xi$  is satisfied for some  $\xi > 0$ . We take  $K \geq \max(4/\xi, 32 \cdot (\zeta + 1)/(\epsilon\xi))$ . We now consider the superposition MRS code with block size  $K$  and parameter  $g(\alpha)$ .

Fix any channel capacity  $c$  with  $r(c) > 0$ . Let  $\alpha_1 = r(c)$ , and  $\alpha_0 = r(c) - \epsilon$ . At time instance  $i$ , the sender generate an encoding symbol from the block  $B_{\lceil i \cdot A_i/K \rceil}$  (note that  $B_i = M_{(i-1)K+1}^{iK}$ ). Let  $F_A(\alpha) = \int_0^{\alpha} g(x)dx - \alpha g(\alpha)$ . The probability that  $B_j$  is chosen is

$$\mathbb{P} \left\{ A_i \in \left( \frac{K(j-1)}{i}, \frac{Kj}{i} \right] \right\} = F_A \left( \frac{Kj}{i} \right) - F_A \left( \frac{K(j-1)}{i} \right).$$

Let  $n \geq \frac{K}{\epsilon}$ , and  $j \leq \frac{n\alpha_0}{K} + 1$ . We will study whether the block  $B_j$  can be decoded using  $Y_1^n$  with error probability less than  $\epsilon$  when the channel capacity is  $c$ . If so, then  $M_1^{\lfloor \alpha_0 n \rfloor}$  can be decoded using  $Y_1^n$  with bit error probability less than  $\epsilon$  whenever  $n \geq \frac{K}{\epsilon}$ , and thus the rate-capacity pair  $(\alpha_1, c)$  is  $\epsilon$ -admissible. Let  $S$  be

the random variable representing the number of times  $B_j$  is chosen in the encoding of  $X_i$ , and  $X_i$  is not erased, for  $i = 1, \dots, n$ . Its expected value is given by

$$\mathbb{E}[S] = c \cdot \sum_{i=1}^n \left( F_A \left( \frac{Kj}{i} \right) - F_A \left( \frac{K(j-1)}{i} \right) \right).$$

Note that  $F_A$  is monotonically increasing and not greater than 1 (and therefore  $F_A \left( \frac{Kj}{nx} \right)$  is monotonically decreasing with respect to  $x$ ), we have

$$\begin{aligned} \sum_{i=1}^n F_A \left( \frac{Kj}{i} \right) &\geq n \left( \int_{1/n}^1 F_A \left( \frac{Kj}{nx} \right) dx \right) \\ &\geq n \left( \int_0^1 F_A \left( \frac{Kj}{nx} \right) dx \right) - 1, \\ \sum_{i=1}^n F_A \left( \frac{K(j-1)}{i} \right) &\leq n \left( \int_0^1 F_A \left( \frac{K(j-1)}{nx} \right) dx \right). \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E}[S] &\geq cn \left( \int_0^1 \left( F_A \left( \frac{Kj}{nx} \right) - F_A \left( \frac{K(j-1)}{nx} \right) \right) dx \right) - c \\ &\geq cn \left( \int_0^1 \left( F_A \left( \frac{Kj}{nx} \right) - F_A \left( \frac{K(j-1)}{nx} \right) \right) dx \right) - 1, \end{aligned}$$

where

$$\begin{aligned} \int_0^1 F_A \left( \frac{\alpha}{x} \right) dx &= F_A(\alpha) - \int_0^1 x dF_A \left( \frac{\alpha}{x} \right) \\ &= F_A(\alpha) + \alpha \int_{\alpha}^{\infty} \frac{1}{x} dF_A(x) \\ &= F_A(\alpha) + \alpha \int_{\alpha}^{\infty} \frac{1}{x} \cdot -x dg(x) \\ &= F_A(\alpha) - \alpha \int_{\alpha}^{\infty} 1 dg(x) \\ &= F_A(\alpha) + \alpha g(\alpha) \\ &= \int_0^{\alpha} g(x) dx. \end{aligned}$$

Hence, as  $g(x)$  is monotonically decreasing,

$$\begin{aligned} \mathbb{E}[S] &\geq cn \cdot \left( \int_0^{Kj/n} g(x) dx - \int_0^{K(j-1)/n} g(x) dx \right) - 1 \\ &= cn \cdot \int_{K(j-1)/n}^{Kj/n} g(x) dx - 1 \\ &\geq cK \cdot g \left( \frac{Kj}{n} \right) - 1 \\ &\geq cK \cdot g \left( \alpha_0 + \frac{K}{n} \right) - 1 \\ &\geq cK \cdot g(\alpha_0 + \epsilon) - 1 \\ &= cK \cdot g(\alpha_1) - 1 \\ &\geq K(1 + \xi) - 1. \end{aligned}$$

On the other hand,

$$\sum_{i=1}^n F_A\left(\frac{Kj}{i}\right) \leq n \left( \int_0^1 F_A\left(\frac{Kj}{nx}\right) dx \right),$$

and

$$\begin{aligned} \sum_{i=1}^n F_A\left(\frac{K(j-1)}{i}\right) &\geq n \left( \int_{1/n}^1 F_A\left(\frac{K(j-1)}{nx}\right) dx \right) \\ &\geq n \left( \int_0^1 F_A\left(\frac{K(j-1)}{nx}\right) dx \right) - 1. \end{aligned}$$

Let

$$p_i = c \left( F_A\left(\frac{Kj}{i}\right) - F_A\left(\frac{K(j-1)}{i}\right) \right).$$

The variance of  $S$  can be given by

$$\begin{aligned} \text{Var}[S] &= \sum_{i=1}^n (p_i - p_i^2) \\ &\leq \sum_{i=1}^n p_i \\ &\leq cn \cdot \left( \int_0^1 \left( F_A\left(\frac{Kj}{nx}\right) - F_A\left(\frac{K(j-1)}{nx}\right) \right) dx \right) + c \\ &\leq n \cdot \left( \int_0^1 \left( F_A\left(\frac{Kj}{nx}\right) - F_A\left(\frac{K(j-1)}{nx}\right) \right) dx \right) + 1 \\ &= n \cdot \left( \int_0^{Kj/n} g_2(x) dx - \int_0^{K(j-1)/n} g_2(x) dx \right) + 1 \\ &= n \cdot \int_{K(j-1)/n}^{Kj/n} g_2(x) dx + 1 \\ &\leq K\zeta + 1, \\ \sigma_S &\leq \sqrt{K} \cdot \sqrt{\zeta + 1}. \end{aligned}$$

As a result, for  $K$  large enough,  $S$  is close to  $\mathbb{E}[S]$  with high probability. Cheby-shev's inequality gives

$$\mathbb{P} \left\{ S < K(1 + \xi) - 1 - \frac{\sqrt{K} \cdot \sqrt{\zeta + 1}}{\sqrt{\epsilon/2}} \right\} \leq \epsilon/2,$$

where

$$\begin{aligned} &K(1 + \xi) - 1 - \frac{\sqrt{K} \cdot \sqrt{\zeta + 1}}{\sqrt{\epsilon/2}} \\ &= K \left( 1 + \xi - \frac{1}{K} - \frac{\sqrt{\zeta + 1}}{\sqrt{K\epsilon/2}} \right) \\ &\geq K(1 + \xi/2) \end{aligned}$$

by the assumption  $K \geq \max(4/\xi, 32 \cdot (\zeta + 1)/(\epsilon\xi))$ . When random linear projections are used, for a fixed probability of error  $\epsilon/2$ , the number of symbols needed to decode a block of  $K$  bits is  $K + o(K)$ , which is smaller than  $K(1 + \xi/2)$  for sufficiently large  $K$ . Therefore the rate function  $r(c)$  is  $\epsilon$ -admissible by the superposition MRS code with parameter  $g(\alpha)$  for  $K$  large enough.  $\square$

The following theorem gives the achievable region for the rate-capacity function, which coincides with the region attained by superposition MRS code.

**Theorem 7.** *A rate-capacity function  $r(c)$  is achievable if and only if*

$$\int_0^1 \frac{1}{c} dr(c) \leq 1.$$

*Proof of achievability.* Assume  $r(c)$  is a rate-capacity function satisfying  $\int_0^1 \frac{1}{c} dr(c) \leq 1$ . We can find  $\eta > 0$  such that  $r(c) = 0$  for  $c \leq \eta$ . Let  $g(\alpha) = 1/r^{-1}(\alpha)$ , where  $r^{-1}(\alpha) = \inf\{c | r(c) \geq \alpha\}$ . Then  $g(\alpha)$  is monotonically decreasing and bounded above by  $1/\eta$ . It is implied by  $\int_0^1 \frac{1}{c} dr(c) \leq 1$  that  $\int_0^\infty g(\alpha) d\alpha \leq 1$ . As  $r^{-1}(r(c)) \leq c$ , we also have  $c \cdot g(r(c)) \geq 1$  for all  $c > 0$  with  $r(c) > 0$ .

Fix  $\epsilon > 0$ . Note that  $\int_0^\infty g(\alpha + \epsilon/2) d\alpha < 1$ . We can define a function  $g_2(\alpha)$  by

$$g_2(\alpha) = (1 + \xi) \cdot g(\alpha + \epsilon/2),$$

where  $\xi > 0$  such that  $\int_0^\infty g_2(\alpha) d\alpha = 1$ . We know  $g_2$  is bounded above by  $(1 + \xi)/\eta$ . Define a new rate-capacity function  $r_2(c) = \max(r(c) - \epsilon/2, 0)$ . We have, for any  $c$  with  $r_2(c) > 0$ ,

$$c \cdot g_2(r_2(c)) = c \cdot (1 + \xi)g(r(c)) \geq 1 + \xi.$$

Consider the superposition MRS code with parameter  $g_2(\alpha)$ . By Theorem 6, the rate-capacity function  $r_2(c)$  is  $\epsilon/2$ -admissible by the code for sufficiently large block size, which implies that  $r(c)$  is  $\epsilon$ -admissible by the code.  $\square$

*Proof of converse.* The proof employs a similar idea as in [2]. As  $r(c)$  is a rate-capacity function, we can find  $\eta > 0$  such that  $r(c) = 0$  for  $c \leq \eta$ . For any  $\epsilon > 0$ , consider a code where  $r(c)$  is  $\epsilon$ -admissible. The message  $\{M_i\}$  are encoded into binary symbols  $\{X_i\}$ , and sent through an erasure channel with capacity  $c$  (we call it Channel  $c$ ) to give  $\{Y_{c,i}\}$  for all  $c > 0$ . Assume we have the following for any  $c$ ,

$$\max_{m \leq n(r(c) - \epsilon)} \mathbb{P} \left\{ M_m \neq \widetilde{M}_m(Q, Y_{c,1}^n) \right\} < \epsilon \text{ for any } n \geq N_0.$$

Let  $N \geq N_0$ . Let  $E_{c,i}$  be the indicator of the events of erasure in Channel  $c$  ( $E_{c,i} = 1$  means that  $Y_{c,i} = e$ ). Note that whether the message can be decoded at the receiver with Channel  $c$  depends only on the marginal distribution of  $\{E_{c,i}, X_i, M_i, Q\}_{i \in \mathbb{N}}$ , and is conditional independent of  $E_{c_2,i}$  for  $c_2 \neq c$  given  $\{E_{c,i}, X_i, M_i, Q\}_{i \in \mathbb{N}}$ . Therefore, we may modify the joint distribution of  $\{E_{c,i}\}_{c \in (0,1], i \in \mathbb{N}}$  without affecting the result as long as the distributions of  $\{E_{c,i}, X_i, M_i, Q\}_{i \in \mathbb{N}}$  are preserved. From now on, we assume the channels are cascaded, i.e. whenever  $c_0 < c_1 \leq 1$ , the Markov chain  $X_i \rightarrow Y_{c_1,i} \rightarrow Y_{c_0,i}$  holds, and

$$E_{c_0,i} = \begin{cases} E_{c_1,i} & \text{w.p. } c_0/c_1 \\ 1 & \text{w.p. } 1 - c_0/c_1. \end{cases}$$

Let

$$(3.2) \quad r_2(c) = \max(r(c) - \epsilon, 0).$$

Define

$$(3.3) \quad f(c) = \frac{1}{Nc} \cdot H(Y_{c,1}^N | Q, M_1^{\lfloor N r_2(c) \rfloor}, E_{c,1}^N).$$



The common random variable  $Q$  will be omitted for simplicity. Consider Channel  $c_0$  and Channel  $c_1$  where  $\eta/2 \leq c_0 < c_1 \leq 1$ . Let  $k_0 = \lfloor Nr_2(c_0) \rfloor$ , and  $k_1 = \lfloor Nr_2(c_1) \rfloor$ . Note that

$$(3.4) \quad \begin{aligned} & H(Y_{c_1,1}^N | M_1^{k_1}, E_{c_1,1}^N) \\ &= H(Y_{c_1,1}^N, M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N) - H(M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N), \end{aligned}$$

where, due to the assumption that  $M_i$  are i.i.d. uniform in  $\{0, 1\}$ ,

$$H(M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N) = H(M_{k_0+1}^{k_1}) = k_1 - k_0 \geq N(r_2(c_1) - r_2(c_0)) - 1.$$

As  $M_1^{k_1}$  can be decoded using  $Y_{c_1,1}^N$  with bit error probability less than  $\epsilon$ , by Fano's inequality,

$$\begin{aligned} & H(Y_{c_1,1}^N, M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N) \\ &= H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) \\ &\quad + H(M_{k_0+1}^{k_1} | Y_{c_1,1}^N, M_1^{k_0}, E_{c_1,1}^N) \\ &\leq H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) \\ &\quad + \sum_{i=k_0+1}^{k_1} H(M_i | Y_{c_1,1}^N, M_1^{k_0}, E_{c_1,1}^N) \\ &\leq H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) + (k_1 - k_0) \cdot H(\epsilon) \\ &\leq H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) + (N(r_2(c_1) - r_2(c_0)) + 1) \cdot H(\epsilon) \\ &\leq H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) + (N(r_2(c_1) - r_2(c_0))) \cdot H(\epsilon) + 2. \end{aligned}$$

Let  $E'_i \stackrel{i.i.d.}{\sim} \text{Bern}(1 - c_0/c_1)$  with

$$E_{c_0,i} = \begin{cases} E_{c_1,i} & \text{if } E'_i = 0 \\ 1 & \text{if } E'_i = 1. \end{cases}$$

We can obtain

$$\begin{aligned}
& H(Y_{c_0,1}^N | M_1^{k_0}, E_{c_0,1}^N) \\
&= \sum_{i=1}^N H(Y_{c_0,i} | M_1^{k_0}, Y_{c_0,1}^{i-1}, E_{c_0,1}^N) \\
&\stackrel{(i)}{\geq} \sum_{i=1}^N H(Y_{c_0,i} | M_1^{k_0}, Y_{c_1,1}^{i-1}, E_{c_0,1}^N) \\
&\stackrel{(ii)}{=} \sum_{i=1}^N H(Y_{c_0,i} | M_1^{k_0}, Y_{c_1,1}^{i-1}, E_{c_0,i}) \\
&\geq \sum_{i=1}^N H(Y_{c_0,i} | M_1^{k_0}, Y_{c_1,1}^{i-1}, E_{c_0,i}, E'_i) \\
&\stackrel{(iii)}{=} \sum_{i=1}^N \left( \frac{c_0}{c_1} \cdot H(Y_{c_0,i} | M_1^{k_0}, Y_{c_1,1}^{i-1}, E_{c_0,i}, E'_i = 0) \right) \\
&= \sum_{i=1}^N \left( \frac{c_0}{c_1} \cdot H(Y_{c_1,i} | M_1^{k_0}, Y_{c_1,1}^{i-1}, E_{c_1,i}, E'_i = 0) \right) \\
&\stackrel{(iv)}{=} \sum_{i=1}^N \left( \frac{c_0}{c_1} \cdot H(Y_{c_1,i} | M_1^{k_0}, Y_{c_1,1}^{i-1}, E_{c_1,i}) \right) \\
&\stackrel{(v)}{\geq} \sum_{i=1}^N \left( \frac{c_0}{c_1} \cdot H(Y_{c_1,i} | M_1^{k_0}, Y_{c_1,1}^{i-1}, E_{c_1,1}^N) \right) \\
&= \frac{c_0}{c_1} \cdot H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N),
\end{aligned}$$

where (i) is due to  $H(Y_{c_0,1}^{i-1} | Y_{c_1,1}^{i-1}, E_{c_0,1}^N) = 0$ , (ii) is due to  $(E_{c_0,1}^{i-1}, E_{c_0,i+1}^N) \perp\!\!\!\perp Y_{c_0,i} | (M_1^{k_0}, Y_{c_1,1}^{i-1})$ , (iii) is obtained by conditioning on  $E'_i$  and by  $Y_{c_0,i} = e$  when  $E'_i = 1$ , (iv) is due to  $E'_i \perp\!\!\!\perp (Y_{c_1,i}, M_1^{k_0}, Y_{c_0,1}^{i-1}, E_{c_1,i})$ , and (v) is due to  $H(E_{c_1,1}^{i-1} | Y_{c_1,1}^{i-1}) = 0$  and  $E_{c_1,i+1}^N \perp\!\!\!\perp Y_{c_1,i} | (M_1^{k_0}, Y_{c_1,1}^{i-1})$ .

Hence by (3.4),

$$\begin{aligned}
& H(Y_{c_1,1}^N | M_1^{k_1}, E_{c_1,1}^N) \\
&\leq \frac{c_1}{c_0} \cdot H(Y_{c_0,1}^N | M_1^{k_0}, E_{c_0,1}^N) + (N(r_2(c_1) - r_2(c_0))) \cdot H(\epsilon) + 2 \\
&\quad - (N(r_2(c_1) - r_2(c_0)) - 1),
\end{aligned}$$

After replacing the terms by  $f(c)$  using (3.3),

$$\frac{1}{c_1} (r_2(c_1) - r_2(c_0)) \leq \frac{f(c_0) - f(c_1) + \frac{3}{Nc_1}}{1 - H(\epsilon)} \leq \frac{f(c_0) - f(c_1) + \frac{6}{N\eta}}{1 - H(\epsilon)}.$$

Due to the monotonicity of  $r_2(c)$ ,

$$\begin{aligned}
& \int_{c_0}^{c_1} \frac{1}{c} dr_2(c) - \left( \frac{1}{c_0} - \frac{1}{c_1} \right) (r_2(c_1) - r_2(c_0)) \\
& \leq \frac{1}{c_0} (r_2(c_1) - r_2(c_0)) - \left( \frac{1}{c_0} - \frac{1}{c_1} \right) (r_2(c_1) - r_2(c_0)) \\
& = \frac{1}{c_1} (r_2(c_1) - r_2(c_0)) \\
& \leq \frac{f(c_0) - f(c_1) + \frac{6}{N\eta}}{1 - H(\epsilon)}.
\end{aligned}$$

Let  $m = \lfloor \sqrt{N} \rfloor$ , and  $c_0 = c'_0 < c'_1 < \dots < c'_m = c_1$  such that  $r_2(c)$  is continuous at  $c = c'_1, \dots, c'_{m-1}$  and

$$\frac{1}{c'_i} - \frac{1}{c'_{i+1}} < \frac{2}{m} \left( \frac{1}{c_0} - \frac{1}{c_1} \right) \text{ for } i = 0, \dots, m-1.$$

We can always find such  $c'_i$  as a monotonic function has at most countable discontinuities. Note that for  $i = 0, \dots, m-1$ ,

$$\frac{1}{c'_i} - \frac{1}{c'_{i+1}} < \frac{2}{m} \left( \frac{1}{c_0} - \frac{1}{c_1} \right) \leq \frac{4}{m\eta}$$

as  $\eta/2 \leq c_0 < c_1$ . Then we have

$$\begin{aligned}
& \int_{c_0}^{c_1} \frac{1}{c} dr_2(c) \\
& = \sum_{i=0}^{m-1} \int_{c'_i}^{c'_{i+1}} \frac{1}{c} dr_2(c) \\
& \leq \sum_{i=0}^{m-1} \left( \frac{f(c'_i) - f(c'_{i+1}) + \frac{6}{N\eta}}{1 - H(\epsilon)} + \left( \frac{1}{c'_i} - \frac{1}{c'_{i+1}} \right) (r_2(c'_{i+1}) - r_2(c'_i)) \right) \\
& \leq \frac{f(c_0) - f(c_1) + \frac{6m}{N\eta}}{1 - H(\epsilon)} + \frac{4}{m\eta} (r_2(c_1) - r_2(c_0)) \\
& \leq \frac{f(c_0) - f(c_1) + \frac{6m}{N\eta}}{1 - H(\epsilon)} + \frac{4}{m\eta} r(1).
\end{aligned}$$

Note that

$$\begin{aligned}
f(c) &= \frac{1}{Nc} \cdot H(Y_{c,1}^N | Q, M_1^{\lfloor N(r(c)-\epsilon) \rfloor}, E_{c,1}^N) \\
&\leq \frac{1}{Nc} \cdot H(Y_{c,1}^N | E_{c,1}^N) \\
&\leq 1.
\end{aligned}$$

Thus we have, for any  $\eta/2 \leq c_0 < c_1 \leq 1$ ,

$$\int_{c_0}^{c_1} \frac{1}{c} dr_2(c) \leq \frac{1 + \frac{6m}{N\eta}}{1 - H(\epsilon)} + \frac{4}{m\eta} r(1).$$

Therefore we can obtain an inequality on  $r(c)$  using (3.2) by

$$\begin{aligned} \int_{c_0}^{c_1} \frac{1}{c} dr(c) &= \int_{c_0}^{c_1} \frac{1}{c} d(r(c) - \epsilon) \\ &\leq \int_{c_0}^{c_1} \frac{1}{c} dr_2(c) + \frac{\epsilon}{\eta} \\ &\leq \frac{1 + \frac{6m}{N\eta}}{1 - H(\epsilon)} + \frac{4}{m\eta} r(1) + \frac{\epsilon}{\eta}. \end{aligned}$$

The inequality holds for arbitrarily large  $N$  and arbitrarily small  $\epsilon$ . We can conclude that

$$\int_0^1 \frac{1}{c} dr(c) = \int_{\eta/2}^1 \frac{1}{c} dr(c) \leq 1.$$

□

*Remark.* It is shown in the theorem that the achievable region of MRS codes coincides with that of priority encoding transmission or multilevel diversity coding. When there are more than one transmitters, the MRS codes no longer admit the same region as priority encoding transmission or multilevel diversity coding in general.

#### 4. THE MULTIPLE TRANSMITTER SETTING

In this section, we will discuss the case where there are  $d$  transmitters which cooperate to send the same sequence of data  $\{M_i\}$ , but they may or may not be transmitting the same sequence of encoding symbols. There are multiple receivers that wish to decode  $\{M_i\}$  sequentially at different rates. A receiver has an erasure channel connected to each of the  $d$  transmitters, and the channels may have different capacities. The central question of this section is that, given a set of receivers with different capacities and different rate requirements, is it possible to design a code which can satisfy the need of all receivers?

We call the transmitters as Transmitter  $k$ , where  $k = 1, \dots, d$ . We denote the symbol sent by Transmitter  $k$  at time  $n$  by  $X_{k,n} \in \{0, 1\}$ . For a receiver with a channel from Transmitter  $k$  with capacity  $c$ , denote the symbol received from Transmitter  $k$  by  $Y_{k,c,n} \in \{0, 1, e\}$ , and the indicator of erasure  $E_{k,c,n} \in \{0, 1\}$  ( $E_{k,c,n} = 1$  indicates an erasure). The definition of a multi-transmitter multi-rate sequential code is similar to that in the single transmitter case, and will be omitted.

For a vector of channel capacities  $\mathbf{c} \in [0, 1]^d$ , where  $c_k$  is the capacity of the channel to Transmitter  $k$ , we write

$$\begin{aligned} X_n &= (X_{1,n}, \dots, X_{d,n}), \\ Y_{\mathbf{c},n} &= (Y_{1,c_1,n}, \dots, Y_{d,c_d,n}), \\ E_{\mathbf{c},n} &= (E_{1,c_1,n}, \dots, E_{d,c_d,n}). \end{aligned}$$

We use the notation  $Y_{\mathbf{c},a}^b = (Y_{\mathbf{c},a}, Y_{\mathbf{c},a+1}, \dots, Y_{\mathbf{c},b})$ . From now on, we refer to the receiver with channel capacities  $\mathbf{c}$  as Receiver  $\mathbf{c}$ . We write the sum of capacities in  $\mathbf{c}$  by  $\Sigma(\mathbf{c}) = \sum_{i=1}^d c_i$ .

**Definition 8** (admissible pair). A rate-capacity pair  $(r, \mathbf{c})$  is called  $\epsilon$ -admissible by a code if Receiver  $\mathbf{c}$  can decode the first  $N(r - \epsilon)$  bits  $M_1^{N(r-\epsilon)}$  with bit error probability less than  $\epsilon$  when the first  $N$  symbols  $Y_{\mathbf{c},1}^N$  are received, for sufficiently large  $N$ . More precisely, there exist  $N_0$  such that

$$\mathbb{P} \left\{ M_m \neq \widetilde{M}_m(Q, Y_{\mathbf{c},1}^N) \right\} \leq \epsilon \text{ for any } N \geq N_0, m \leq N(r - \epsilon).$$

We use a function to characterize the rate of a code. We call  $r : [0, 1]^d \rightarrow [0, \infty)$  a *rate-capacity function* if it is monotonically increasing and right continuous along each of the  $d$  dimensions, and there exist an  $\eta > 0$  such that  $r(\mathbf{c}) = 0$  for  $\Sigma(\mathbf{c}) \leq \eta$ .

**Definition 9** (rate of MRS code). A rate-capacity function  $r(\mathbf{c})$  is called  $\epsilon$ -admissible by a code if all of the pairs  $(\mathbf{c}, r(\mathbf{c}))$  are  $\epsilon$ -admissible by the code.

**Definition 10** (achievable rate-capacity functions). A rate-capacity function  $r(\mathbf{c})$  is *achievable* if for any  $\epsilon > 0$ , there exist a code where  $r(\mathbf{c})$  is  $\epsilon$ -admissible by that code.

The superposition MRS code for multiple transmitters is similar to that for single transmitter.

**Definition 11** (superposition MRS code). A superposition MRS code is characterized by the block size  $K$  and the parameter  $g : [0, \infty) \rightarrow [0, \infty)^d$  which is bounded, monotonically decreasing and left continuous along each dimension with  $\int_0^\infty g_k(\alpha) d\alpha = 1$  for  $k = 1, \dots, d$  (write  $g_k(\alpha)$  for the  $k$ -th entry of  $g(\alpha)$ ). Transmitter  $k$  generates encoding symbols using the single transmitter superposition MRS code with block size  $K$  and parameter  $g_k(\alpha)$ .

We say that a rate-capacity function is achievable by superposition MRS code if it is  $\epsilon$ -admissible by a superposition MRS code for arbitrarily small  $\epsilon$ . We give the necessary and sufficient condition on the achievability by superposition MRS code.

**Theorem 12.** *The rate-capacity function  $r(\mathbf{c})$  is achievable by superposition MRS code if and only if there exists a function  $g : [0, \infty) \rightarrow [0, \infty)^d$  which is bounded, monotonically decreasing and left continuous along each of the  $d$  dimensions satisfying*

$$\int_0^\infty g_k(\alpha) d\alpha = 1 \text{ for } k = 1, \dots, d, \text{ and}$$

$$\mathbf{c} \cdot g(r(\mathbf{c})) \geq 1 \text{ for any } \mathbf{c} \in [0, 1]^d \text{ with } r(\mathbf{c}) > 0.$$

*Proof.* The “if” part is similar to the proof of achievability in Theorem 7. Fix any  $\epsilon > 0$ . Let  $\eta > 0$  such that  $r(\mathbf{c}) = 0$  for  $\Sigma(\mathbf{c}) \leq \eta$ . Define  $g' : [0, \infty) \rightarrow [0, \infty)^d$  by

$$g'_k(\alpha) = (1 + \xi_k) \cdot g_k(\alpha + \epsilon/2),$$

where  $\xi_k > 0$  such that  $\int_0^\infty g'_k(\alpha) d\alpha = 1$ . Define a new rate-capacity function  $r_2(\mathbf{c}) = \max(r(\mathbf{c}) - \epsilon/2, 0)$ . We have, for any  $\mathbf{c}$  with  $r_2(\mathbf{c}) > 0$ ,

$$\mathbf{c} \cdot g'(r_2(\mathbf{c})) = \mathbf{c} \cdot (1 + \xi) g'(r_2(\mathbf{c})) \geq 1 + \xi.$$

Consider the multiple transmitter superposition MRS code with parameter  $g(\alpha)$ . Applying Theorem 6 on each dimension, the rate-capacity function  $r_2(\mathbf{c})$  is  $\epsilon/2$ -admissible by the code for sufficiently large block size, which implies that  $r(\mathbf{c})$  is  $\epsilon$ -admissible by the code.

For the “only if” part, let  $\epsilon > 0$  and consider a superposition MRS code with block size  $K$  and parameter  $g(\alpha)$  in which  $r(\mathbf{c})$  is  $\epsilon$ -admissible. Consider the receiver with capacities  $\mathbf{c}$  which decode at rate  $r(\mathbf{c}) - \epsilon$ . Using similar arguments as in 6, fixing any block with sufficiently large index, the expected number of times when a received symbol from Transmitter  $k$  is encoded from the block can be given by  $c_k K \cdot g_k(r(\mathbf{c}) - \epsilon)$ . As at least  $K$  received symbols is required to decode the block, we have

$$\mathbf{c} \cdot g(r(\mathbf{c}) - \epsilon) = \sum_k c_k g_k(r(\mathbf{c}) - \epsilon) \geq 1.$$

As  $g(\alpha)$  is left continuous, the proof can be completed by taking  $\epsilon \rightarrow 0$ .  $\square$

## 5. GENERAL NON-OPTIMALITY OF SUPERPOSITION CODES

We have shown in Section 3 that superposition codes are optimal for single transmitter. However, in the multiple transmitter setting, the superposition MRS codes are not optimal in general. We will provide a counter example.

**Example 13.** Consider the two transmitter case. Given the block size  $K$  and  $0 < r < 2/3$ , the code is constructed by the following process. For Transmitter  $k$  ( $k = 1, 2$ ), at time  $i$ , we generate  $X_{k,i}$  by taking a random linear combination of the bits in the blocks  $M'_{2, \lceil i/(4K) \rceil - 2 + k}$ ,  $M'_{4, \lceil i/(4K) \rceil - 1 - k}$ , and  $M'_{4, \lceil i/(4K) \rceil + 1 - k}$ , where  $M'_i = M_{(i-1)K+1}^{iK}$ .

Consider Transmitter 1, which generates  $4K$  encoding bits from the three blocks  $M'_{2, \lceil i/(4K) \rceil - 1}$ ,  $M'_{4, \lceil i/(4K) \rceil - 2}$ , and  $M'_{4, \lceil i/(4K) \rceil}$ , with a total of  $3K$  bits. Hence, a receiver which can only receive from Transmitter 1 with channel capacity  $3/4$  can decode the three blocks. When  $\lceil i/(4K) \rceil = n$ , the blocks  $M'_{2n-1}$ ,  $M'_{4n-2}$ , and  $M'_{4n}$  can be decoded, which covers all the blocks. The rate-capacity pair  $(\frac{1}{2}, (\frac{3}{4}, 0))$  is admissible. Similar for  $(\frac{1}{2}, (0, \frac{3}{4}))$ .

Consider a receiver which receives from Transmitter 1 and 2, each with channel capacity  $1/2$ . It receives  $2K$  bits encoding the three blocks  $M'_{2, \lceil i/(4K) \rceil - 1}$ ,  $M'_{4, \lceil i/(4K) \rceil - 2}$ ,  $M'_{4, \lceil i/(4K) \rceil}$ , and also  $2K$  bits encoding the three blocks  $M'_{2, \lceil i/(4K) \rceil}$ ,  $M'_{4, \lceil i/(4K) \rceil - 3}$ ,  $M'_{4, \lceil i/(4K) \rceil - 1}$ . When  $n \stackrel{\text{def}}{=} \lceil i/(4K) \rceil = 1$ , the  $4K$  bits encoding the blocks  $M'_1$  to  $M'_4$  are sufficient to decode the blocks. When  $n \geq 2$ , assume the blocks  $M'_m$  for  $m \leq 4n - 4$  are already decoded, then there are  $2K$  bits encoding the two blocks  $M'_{4n-2}$  and  $M'_{4n}$  ( $M'_{2n-1}$  is already decoded) which are sufficient to decode the blocks, and  $2K$  bits encoding the two blocks  $M'_{4n-3}$  and  $M'_{4n-1}$  which are sufficient to decode the blocks. The rate-capacity pair  $(1, (\frac{1}{2}, \frac{1}{2}))$  is admissible.

However, the rate-capacity pairs  $(\frac{1}{2}, (\frac{3}{4}, 0))$  and  $(1, (\frac{1}{2}, \frac{1}{2}))$  cannot be simultaneously achieved by superposition MRS code. If it can be achieved by superposition MRS code, by Theorem 12, there is a monotonically decreasing function  $g : [0, \infty) \mapsto [0, \infty)^2$  satisfying  $\int_0^\infty g_1(\alpha) d\alpha = \int_0^\infty g_2(\alpha) d\alpha = 1$  and  $\frac{3}{4}g_1(\frac{1}{2}) \geq 1$ ,  $\frac{1}{2}g_1(1) + \frac{1}{2}g_2(1) \geq 1$ . As  $g_1(1), g_2(1) \leq 1$ , we have  $g_1(1) = g_2(1) = 1$ , and  $g_1(\alpha) = g_2(\alpha) = 1$  when  $\alpha \leq 1$ , which contradicts with  $\frac{3}{4}g_1(\frac{1}{2}) \geq 1$ .

In the following sections, we will study some special cases in which superposition codes are optimal.

## 6. SOME USEFUL TOOLS

We will present some tools which are used to find the admissible region in certain special cases.

For the sake of simplicity, we write

$$\begin{aligned} J_{N\alpha}^\infty(Y_{\mathbf{c},1}^N) &= H(Y_{\mathbf{c},1}^N | M_1^{\lfloor N\alpha \rfloor}, E_{\mathbf{c},1}^N, Q), \text{ and} \\ J_{N\alpha}^{N\beta}(Y_{\mathbf{c},1}^N) &= I(M_{\lfloor N\alpha \rfloor + 1}^{\lfloor N\beta \rfloor}; Y_{\mathbf{c},1}^N | M_1^{\lfloor N\alpha \rfloor}, E_{\mathbf{c},1}^N, Q). \end{aligned}$$

We use the infinity sign “ $\infty$ ” in  $J_{N\alpha}^\infty(Y_{\mathbf{c},1}^N)$  as the encoding symbols  $X_i$  are encoded from  $M_1^\infty$  and  $Q$ , and  $Y_{\mathbf{c},i}$  can be determined by  $X_i$  and  $E_{\mathbf{c},i}$ . As a result,

$$H(Y_{\mathbf{c},1}^N | M_1^\infty, E_{\mathbf{c},1}^N, Q) = 0,$$

and therefore

$$\begin{aligned} J_{N\alpha}^\infty(Y_{\mathbf{c},1}^N) &= H(Y_{\mathbf{c},1}^N | M_1^{\lfloor N\alpha \rfloor}, E_{\mathbf{c},1}^N, Q) \\ &= I(M_{\lfloor N\alpha \rfloor + 1}^\infty; Y_{\mathbf{c},1}^N | M_1^{\lfloor N\alpha \rfloor}, E_{\mathbf{c},1}^N, Q). \end{aligned}$$

(Note that the support of  $M_1^\infty$  is uncountable. The above equations only serve as the intuition behind the definition.)

The quantity  $J_{N\alpha}^{N\beta}(Y_{\mathbf{c},1}^N)$  roughly corresponds to the amount of information in the first  $N$  encoding symbols dedicated to encode the interval of data  $M_{[N\alpha]+1}^{[N\beta]}$ . Furthermore, we define

$$\begin{aligned}\bar{J}_\alpha^\infty(Y_{\mathbf{c}}, T) &= \frac{1}{T} \cdot \int_0^T J_{\alpha \lfloor e^x \rfloor}^\infty(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) \cdot e^{-x} dx, \\ \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T) &= \frac{1}{T} \cdot \int_0^T J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) \cdot e^{-x} dx.\end{aligned}$$

By observing  $H(Y_{\mathbf{c},1}^N | E_{\mathbf{c},1}^N) \leq N\Sigma(\mathbf{c})$ , we know the limits are finite as

$$\begin{aligned}\frac{1}{T} \cdot \int_0^T J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) \cdot e^{-x} dx \\ \leq \frac{1}{T} \cdot \int_0^T \Sigma(\mathbf{c}) \lfloor e^x \rfloor \cdot e^{-x} dx \\ \leq \Sigma(\mathbf{c}),\end{aligned}$$

and thus

$$(6.1) \quad \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T) \leq \Sigma(\mathbf{c}).$$

Similarly, by considering  $J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) \leq (\beta - \alpha) \lfloor e^x \rfloor$ , we can obtain

$$(6.2) \quad \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T) \leq \beta - \alpha.$$

If  $(\mathbf{c}, r)$  is  $\epsilon$ -admissible, then whenever  $r \geq \beta > \alpha \geq 0$ ,  $M_{[N\alpha]+1}^{[N(\beta-\epsilon)]}$  can be decoded using  $Y_{\mathbf{c},1}^N$  for sufficiently large  $N$ . By Fano's inequality (note that the case where  $\beta - \epsilon \leq \alpha$  is obvious),

$$\begin{aligned}\liminf_{T \rightarrow \infty} \frac{1}{T} \cdot \int_0^T J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) \cdot e^{-x} dx \\ \geq \liminf_{T \rightarrow \infty} \frac{1}{T} \cdot \int_0^T (\beta - \alpha - \epsilon)(1 - H(\epsilon)) \lfloor e^x \rfloor \cdot e^{-x} dx \\ = (\beta - \alpha - \epsilon)(1 - H(\epsilon)),\end{aligned}$$

and therefore

$$(6.3) \quad \liminf_{T \rightarrow \infty} \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T) \geq (\beta - \alpha - \epsilon)(1 - H(\epsilon)).$$

Also it is clear that for  $\gamma < \alpha < \beta$ ,

$$(6.4) \quad \bar{J}_\gamma^\beta(Y_{\mathbf{c}}, T) = \bar{J}_\gamma^\alpha(Y_{\mathbf{c}}, T) + \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T)$$

The following lemma can be readily observed.

**Lemma 14.** For  $\mathbf{c} = \mathbf{c}_1 + \mathbf{c}_2$ ,  $\mathbf{c}_1, \mathbf{c}_2 \geq 0$ , we have

$$\bar{J}_\alpha^\beta(Y_{\mathbf{c}_1}, T) \leq \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T) \leq \bar{J}_\alpha^\beta(Y_{\mathbf{c}_1}, T) + \bar{J}_\alpha^\infty(Y_{\mathbf{c}_2}, T).$$

*Proof.* As the quantities depend only on the marginal distributions of  $\{E_{\mathbf{c},i}, X_i, M_i, Q\}$ ,  $\{E_{\mathbf{c}_1,i}, X_i, M_i, Q\}$  and  $\{E_{\mathbf{c}_2,i}, X_i, M_i, Q\}$ , but not the joint distribution between  $E_{\mathbf{c},i}$  and  $E_{\mathbf{c}_1,i}$  and so on. For the purpose of analysis, we assume the non-erasure positions of Receiver  $\mathbf{c}_1$  do not overlap with those of Receiver  $\mathbf{c}_2$ , and erasure happens in Receiver  $\mathbf{c}$  if and only if erasure happens in both Receiver  $\mathbf{c}_1$  and Receiver  $\mathbf{c}_2$ . Then  $Y_{\mathbf{c},1}^N$  has the same information as  $(Y_{\mathbf{c}_1,1}^N, Y_{\mathbf{c}_2,1}^N)$ . We can deduce from  $I(X; Z) \leq I(X, Y; Z) \leq I(X; Z) + H(Y)$  that

$$J_{N\alpha}^{N\beta}(Y_{\mathbf{c}_1,1}^N) \leq J_{N\alpha}^{N\beta}(Y_{\mathbf{c},1}^N) \leq J_{N\alpha}^{N\beta}(Y_{\mathbf{c}_1,1}^N) + J_{N\alpha}^\infty(Y_{\mathbf{c}_2,1}^N).$$

The result follows.  $\square$

As the message is transmitted in a sequential manner, the received symbols  $Y_{\mathbf{c},1}^N$  should contain more information about the older messages (the  $M_i$  with smaller  $i$ ) than newer messages, and therefore the average information  $(\beta - \alpha)^{-1} \cdot \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T)$  should increase when  $\alpha$  decrease. This property is proved in the following lemma.

**Lemma 15.** *Let  $0 \leq \gamma \leq \alpha < \beta$ . We have*

$$\liminf_{T \rightarrow \infty} \left( \frac{1}{\beta - \gamma} \cdot \bar{J}_\gamma^\beta(Y_{\mathbf{c}}, T) - \frac{1}{\beta - \alpha} \cdot \bar{J}_\alpha^\beta(Y_{\mathbf{c}}, T) \right) \geq 0.$$

*Proof.* Consider

$$\begin{aligned} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) &= J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) + J_{\gamma \lfloor e^x \rfloor}^{\alpha \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) \\ &\geq J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) + J_{\gamma e^x}^{\alpha e^x}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) - \alpha. \end{aligned}$$

After integrating the second term, we get, for any  $x_0$ ,

$$\begin{aligned} &\int_{x_0}^{x_0 + \log(\beta/\alpha)} J_{\gamma e^x}^{\alpha e^x}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx \\ &\geq \int_{x_0}^{x_0 + \log(\beta/\alpha)} J_{\gamma e^x}^{\alpha e^x}(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx \\ &= \int_{x_0}^{x_0 + \log(\beta/\alpha)} J_{\gamma e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx - \int_{x_0}^{x_0 + \log(\beta/\alpha)} J_{\alpha e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx \\ &= \int_{x_0 - \log(\alpha/\gamma)}^{x_0 - \log(\alpha/\gamma) + \log(\beta/\alpha)} J_{\gamma e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx - \int_{x_0}^{x_0 + \log(\beta/\alpha)} J_{\alpha e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx \\ &= \int_{x_0 - \log(\alpha/\gamma)}^{x_0 + \log(\beta/\alpha^2)} J_{\alpha e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx - \int_{x_0}^{x_0 + \log(\beta/\alpha)} J_{\alpha e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx \\ &= \int_{x_0 - \log(\alpha/\gamma)}^{x_0} J_{\alpha e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx - \int_{x_0 + \log(\beta/\alpha^2)}^{x_0 + \log(\beta/\alpha)} J_{\alpha e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx \\ &= \int_{x_0 - \log(\alpha/\gamma)}^{x_0} J_{\alpha e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx - \int_{x_0 - \log(\alpha/\gamma)}^{x_0} J_{\beta e^x}^\infty(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx \\ &= \int_{x_0 - \log(\alpha/\gamma)}^{x_0} J_{\alpha e^x}^{\beta e^x}(Y_{\mathbf{c},1}^{\lfloor e^{x_0} \rfloor}) dx \\ &\geq \int_{x_0 - \log(\alpha/\gamma)}^{x_0} J_{\alpha e^x}^{\beta e^x}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx \\ &\geq \int_{x_0 - \log(\alpha/\gamma)}^{x_0} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx - \alpha \log \frac{\alpha}{\gamma}. \end{aligned}$$

As a result,

$$\begin{aligned} &\int_{x_0 - \log(\beta/\alpha)}^{x_0} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx \\ &\geq \int_{x_0 - \log(\beta/\alpha)}^{x_0} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) + \int_{x_0 - \log(\beta/\alpha)}^{x_0} J_{\gamma e^x}^{\alpha e^x}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) - \alpha \log \frac{\beta}{\alpha} \\ &\geq \int_{x_0 - \log(\beta/\alpha)}^{x_0} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) + \int_{x_0 - \log(\beta/\gamma)}^{x_0 - \log(\beta/\alpha)} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx - \alpha \log \frac{\beta}{\gamma} \\ (6.5) \quad &\geq \int_{x_0 - \log(\beta/\gamma)}^{x_0} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor}(Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) - \alpha \log \frac{\beta}{\gamma}. \end{aligned}$$



Fix any  $T > 0$ . For each term in (6.5), multiply it with  $e^{-x_0}$  and integrate it from 0 to  $T$ , we get

$$\begin{aligned}
& \int_0^T e^{-x_0} \int_{x_0 - \log(\beta/\alpha)}^{x_0} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx dx_0 \\
&= \int_0^T \int_x^{\min(T, x + \log(\beta/\alpha))} e^{-x_0} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx_0 dx \\
&\leq \int_0^T \int_x^{x + \log(\beta/\alpha)} e^{-x_0} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx_0 dx \\
&= \left(1 - \frac{\alpha}{\beta}\right) \int_0^T e^{-x} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx.
\end{aligned}$$

And also

$$\begin{aligned}
& \int_0^T e^{-x_0} \int_{x_0 - \log(\beta/\gamma)}^{x_0} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx dx_0 \\
&= \int_0^T \int_x^{\min(T, x + \log(\beta/\gamma))} e^{-x_0} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx_0 dx \\
&\geq \int_0^{T - \log(\beta/\gamma)} \int_x^{x + \log(\beta/\gamma)} e^{-x_0} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx_0 dx \\
&= \left(1 - \frac{\gamma}{\beta}\right) \int_0^{T - \log(\beta/\gamma)} e^{-x} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx \\
&\geq \left(1 - \frac{\gamma}{\beta}\right) \int_0^T e^{-x} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx - \left(1 - \frac{\gamma}{\beta}\right) (\beta - \alpha) \log \frac{\beta}{\gamma}.
\end{aligned}$$

Therefore, by (6.5),

$$\begin{aligned}
& \left(1 - \frac{\alpha}{\beta}\right) \int_0^T e^{-x} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx \\
&\geq \left(1 - \frac{\gamma}{\beta}\right) \int_0^T e^{-x} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx - \left(1 - \frac{\gamma}{\beta}\right) (\beta - \alpha) \log \frac{\beta}{\gamma} - \int_0^T e^{-x_0} \alpha \log \frac{\beta}{\gamma} dx_0 \\
&= \left(1 - \frac{\gamma}{\beta}\right) \int_0^T e^{-x} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx - \left(1 - \frac{\gamma}{\beta}\right) (\beta - \alpha) \log \frac{\beta}{\gamma} - (1 - e^{-T}) \alpha \log \frac{\beta}{\gamma} \\
&\geq \left(1 - \frac{\gamma}{\beta}\right) \int_0^T e^{-x} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx - \left(\frac{(\beta - \alpha)(\beta - \gamma)}{\beta} + \alpha\right) \cdot \log \frac{\beta}{\gamma}
\end{aligned}$$

Multiply  $\beta / (T \cdot (\beta - \alpha)(\beta - \gamma))$  to both sides, we can obtain

$$\begin{aligned}
& \frac{1}{\beta - \gamma} \cdot \frac{1}{T} \cdot \int_0^T e^{-x} J_{\gamma \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx \\
&\geq \frac{1}{\beta - \alpha} \cdot \frac{1}{T} \cdot \int_0^T e^{-x} J_{\alpha \lfloor e^x \rfloor}^{\beta \lfloor e^x \rfloor} (Y_{\mathbf{c},1}^{\lfloor e^x \rfloor}) dx - \frac{1}{T} \cdot \left(1 + \frac{\alpha\beta}{(\beta - \alpha)(\beta - \gamma)}\right) \cdot \log \frac{\beta}{\gamma}.
\end{aligned}$$

Note that the second term vanishes when  $T \rightarrow \infty$ . The result follows.  $\square$

We now proceed to prove an inequality on achievable rate-capacity functions.

**Lemma 16.** *If a rate-capacity function  $r(\mathbf{c})$  is achievable, then for any  $\mathbf{c} = \mathbf{c}_1 + \mathbf{c}_2 + \dots + \mathbf{c}_n$ , where  $\mathbf{c}_k \geq 0$  and  $r(\mathbf{c}) > r(\mathbf{c}_k)$  for  $k = 1, \dots, n$ , we have*

$$\sum_{k=1}^n \frac{\Sigma(\mathbf{c}_k) - r(\mathbf{c}_k)}{r(\mathbf{c}) - r(\mathbf{c}_k)} \geq 1.$$

*Proof.* Without loss of generality, assume  $r(\mathbf{c}_1) \leq \dots \leq r(\mathbf{c}_n) < r(\mathbf{c})$ . Let  $\mathbf{c}_0 = 0$ . Consider a code in which  $r(\mathbf{c})$  is  $\epsilon$ -admissible. Fix any  $k \in \{1, \dots, n\}$ . By (6.1), for any  $T \geq 0$ ,

$$\bar{J}_0^\infty(Y_{\mathbf{c}_k}, T) \leq \Sigma(\mathbf{c}_k).$$

By (6.4),

$$\begin{aligned} \bar{J}_{r(\mathbf{c}_k)}^\infty(Y_{\mathbf{c}_k}, T) &= \bar{J}_0^\infty(Y_{\mathbf{c}_k}, T) - \bar{J}_0^{r(\mathbf{c}_k)}(Y_{\mathbf{c}_k}, T) \\ &\leq \Sigma(\mathbf{c}_k) - \bar{J}_0^{r(\mathbf{c}_k)}(Y_{\mathbf{c}_k}, T). \end{aligned}$$

Invoking Lemma 14, we obtain

$$\begin{aligned} \bar{J}_{r(\mathbf{c}_k)}^{r(\mathbf{c})}(Y_{\mathbf{c}_1+\dots+\mathbf{c}_{k-1}}, T) &\geq \bar{J}_{r(\mathbf{c}_k)}^{r(\mathbf{c})}(Y_{\mathbf{c}_1+\dots+\mathbf{c}_k}, T) - \bar{J}_{r(\mathbf{c}_k)}^\infty(Y_{\mathbf{c}_k}, T) \\ &\geq \bar{J}_{r(\mathbf{c}_k)}^{r(\mathbf{c})}(Y_{\mathbf{c}_1+\dots+\mathbf{c}_k}, T) - \Sigma(\mathbf{c}_k) + \bar{J}_0^{r(\mathbf{c}_k)}(Y_{\mathbf{c}_k}, T), \end{aligned}$$

It can be deduced using Lemma 15 that

$$\begin{aligned} \liminf_{T \rightarrow \infty} \left( \frac{1}{r(\mathbf{c}) - r(\mathbf{c}_{k-1})} \bar{J}_{r(\mathbf{c}_{k-1})}^{r(\mathbf{c})}(Y_{\mathbf{c}_1+\dots+\mathbf{c}_{k-1}}, T) \right. \\ \left. - \frac{1}{r(\mathbf{c}) - r(\mathbf{c}_k)} \bar{J}_{r(\mathbf{c}_k)}^{r(\mathbf{c})}(Y_{\mathbf{c}_1+\dots+\mathbf{c}_{k-1}}, T) \right) \geq 0, \end{aligned}$$

and therefore

$$\begin{aligned} \liminf_{T \rightarrow \infty} \left( \frac{1}{r(\mathbf{c}) - r(\mathbf{c}_{k-1})} \bar{J}_{r(\mathbf{c}_{k-1})}^{r(\mathbf{c})}(Y_{\mathbf{c}_1+\dots+\mathbf{c}_{k-1}}, T) \right. \\ \left. - \frac{1}{r(\mathbf{c}) - r(\mathbf{c}_k)} \bar{J}_{r(\mathbf{c}_k)}^{r(\mathbf{c})}(Y_{\mathbf{c}_1+\dots+\mathbf{c}_k}, T) + \frac{\Sigma(\mathbf{c}_k) - \bar{J}_0^{r(\mathbf{c}_k)}(Y_{\mathbf{c}_k}, T)}{r(\mathbf{c}) - r(\mathbf{c}_k)} \right) \geq 0, \end{aligned}$$

Summing through  $k = 1, \dots, n$ ,

$$\begin{aligned} \liminf_{T \rightarrow \infty} \left( \frac{1}{r(\mathbf{c}) - r(0)} \bar{J}_{r(0)}^{r(\mathbf{c})}(Y_{\mathbf{c}_0}, T) \right. \\ \left. - \frac{1}{r(\mathbf{c}) - r(\mathbf{c}_n)} \bar{J}_{r(\mathbf{c}_n)}^{r(\mathbf{c})}(Y_{\mathbf{c}}, T) + \sum_{k=1}^n \frac{\Sigma(\mathbf{c}_k) - \bar{J}_0^{r(\mathbf{c}_k)}(Y_{\mathbf{c}_k}, T)}{r(\mathbf{c}) - r(\mathbf{c}_k)} \right) \geq 0, \end{aligned}$$

and thus

$$\begin{aligned} \liminf_{T \rightarrow \infty} \left( \sum_{k=1}^n \frac{\Sigma(\mathbf{c}_k) - \bar{J}_0^{r(\mathbf{c}_k)}(Y_{\mathbf{c}_k}, T)}{r(\mathbf{c}) - r(\mathbf{c}_k)} \right. \\ \left. - \frac{1}{r(\mathbf{c}) - r(\mathbf{c}_n)} \bar{J}_{r(\mathbf{c}_n)}^{r(\mathbf{c})}(Y_{\mathbf{c}}, T) \right) \geq 0, \end{aligned}$$

Using (6.3), we have

$$\liminf_{T \rightarrow \infty} \bar{J}_0^{r(\mathbf{c}_k)}(Y_{\mathbf{c}_k}, T) \geq (r(\mathbf{c}_k) - \epsilon) (1 - H(\epsilon)),$$

and

$$\liminf_{T \rightarrow \infty} \bar{J}_{r(\mathbf{c}_n)}^{r(\mathbf{c})}(Y_{\mathbf{c}}, T) \geq (r(\mathbf{c}) - r(\mathbf{c}_n) - \epsilon) (1 - H(\epsilon)).$$

Hence,

$$\sum_{k=1}^n \frac{\Sigma(\mathbf{c}_k) - (r(\mathbf{c}_k) - \epsilon)(1 - H(\epsilon))}{r(\mathbf{c}) - r(\mathbf{c}_k)} - \frac{(r(\mathbf{c}) - r(\mathbf{c}_n) - \epsilon)(1 - H(\epsilon))}{r(\mathbf{c}) - r(\mathbf{c}_n)} \geq 0.$$

Let  $\epsilon \rightarrow 0$ . We obtained the desired result as

$$\sum_{k=1}^n \frac{\Sigma(\mathbf{c}_k) - r(\mathbf{c}_k)}{r(\mathbf{c}) - r(\mathbf{c}_k)} \geq 1.$$

□

We will discuss some cases in which superposition coding is provably optimal in the next section.

## 7. ON-OFF MULTICAST NETWORKS

In this section, we consider networks in which there are  $d$  transmitters and  $2^d - 1$  receivers, each having a different set of transmitters to which it is connected. Transmitter  $k$  broadcasts the same information to the receivers it is connected to at rate  $w_k$ . Each receiver has to decode the information at a different rate. There is no erasure in the network. We would like to formulate the criteria on the decoding rates of the receivers in which sequential data transmission is possible.

We first convert the problem into the multi-transmitter MRS setting. Without loss of generality, we assume  $w_k \leq 1$  for  $k = 1, \dots, d$ . We may replace a connection with rate  $w$  by an erasure channel with capacity  $w$ . Note that the decoding requirement does not depend on the joint distribution of erasure events of different receivers. Therefore, the problem can be translated to  $d$ -transmitter MRS. We confine our study to the capacity vectors  $\mathbf{c} \in \{0, w_1\} \times \dots \times \{0, w_d\}$ .

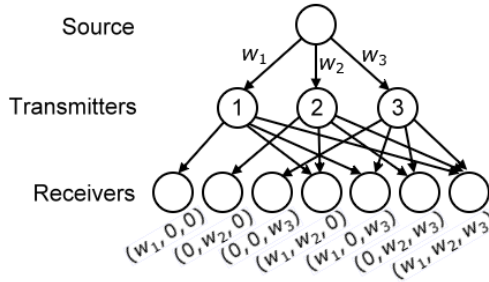


FIGURE 7.1. A three-transmitter on-off multicast network

**7.1. Non-optimality of superposition coding in general.** We can construct a network in which superposition coding is not optimal, using a similar idea as in Example 13.

**Example 17.** The network contains 3 transmitters and 3 receivers, where Transmitter 1 and 2 broadcast at rate 1, and Transmitter 3 broadcast at rate 2. Receiver 1 is connected to Transmitter 1 and 2. Receiver 2 is connected to Transmitter 1 and 3. Receiver 3 is connected to Transmitter 2 and 3. Non-superposition code can achieve the decoding rate  $3/2$  for Receiver 1, and 3 for Receiver 2 and 3, which are not achievable using superposition codes. Please refer to Subsection 7.3 for the proof of achievability and further discussions.

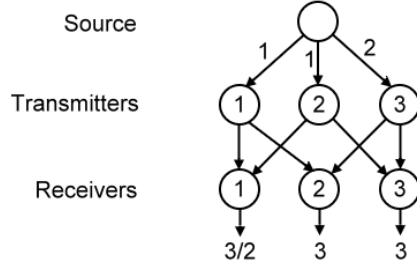


FIGURE 7.2. The network in which superposition codes are not optimal

Although superposition coding is not optimal for networks with 3 transmitters in general, it is optimal for networks with 2 transmitters, which will be shown in the following example.

### 7.2. One-or-all on-off multicast network.

**Example 18** (One-or-all on-off multicast network). There are  $d$  transmitters and  $d + 1$  receivers (numbered  $0, \dots, d$ ), where Transmitter  $k$  broadcasts the same information to Receiver 0 and  $k$  at rate of  $w_k$  bit/s for  $k = 1, \dots, d$ . Receiver  $k$  has to decode the data at rate  $r_k$  bit/s for  $k = 0, \dots, d$ . We would like to characterize the achievable region of  $\{r_k\}$  in terms of  $\{w_k\}$ . Figure 7.3 shows the network when  $d = 2$ .

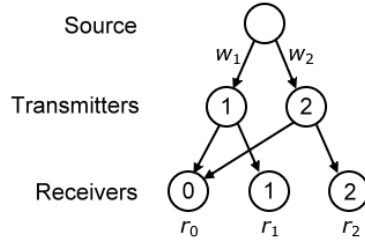


FIGURE 7.3. The two-transmitter one-or-all on-off multicast network

The following theorem shows that superposition coding is optimal.

**Theorem 19.** *Superposition coding is optimal in the one-or-all on-off multicast network, which has an achievable region*

$$r_0 \geq 0, 0 \leq r_k \leq w_k \text{ for } k = 1, \dots, d,$$

$$\text{either } r_0 \leq \max(w_k) \text{ or } \sum_{k=1}^d \frac{w_k - r_k}{r_0 - r_k} \geq 1.$$

*Proof.* The case where  $r_0 \leq \max(w_k)$  is trivial. We assume  $r_0 > w_k$  for all  $k$ . Without loss of generality, assume  $r_1 \leq \dots \leq r_d$ , then we have  $r_1 \leq \dots \leq r_d \leq w_d < r_0$ .

Among the inequalities in the proposed achievable region,  $r_k \leq w_k$  is obvious, and  $\sum_{k=1}^d \frac{w_k - r_k}{r_0 - r_k} \geq 1$  is due to Lemma 16. The converse follows.

We now show superposition MRS code can achieve the region. Assume  $\sum_{k=1}^d \frac{w_k - r_k}{r_0 - r_k} \geq 1$ . Consider the parameter

$$g_k(\alpha) = \begin{cases} \frac{1}{w_k} & \text{when } \alpha \leq r_k \\ \frac{w_k - r_k}{w_k(r_0 - r_k)} & \text{when } r_k < \alpha \leq r_0 \\ 0 & \text{when } \alpha > r_0. \end{cases}$$

Note that

$$\frac{w_k - r_k}{w_k(r_0 - r_k)} \leq \frac{1}{w_k}$$

due to  $r_0 \geq w_k$ . Therefore  $g(\alpha)$  is monotonically decreasing along each dimension. It can be easily checked that  $\int_0^\infty g_k(\alpha) d\alpha = 1$ . It is left to check  $\mathbf{c} \cdot g(r(\mathbf{c})) \geq 1$ .

$$w_k \cdot g_k(r_k) = w_k \cdot \frac{1}{w_k} = 1,$$

$$\begin{aligned} & \sum_{k=1}^d w_k \cdot g_k(r_0) \\ &= \sum_{k=1}^d w_k \cdot \frac{w_k - r_k}{w_k(r_0 - r_k)} \\ &= \sum_{k=1}^d \frac{w_k - r_k}{r_0 - r_k} \geq 1. \end{aligned}$$

By Theorem 12, the region is achievable by superposition MRS code.  $\square$

*Remark.* The achievable region in the MRS setting is different from that in multi-level diversity coding, which admits the larger region

$$r_0 \geq 0, 0 \leq r_k \leq w_k \text{ for } k = 1, \dots, d, r_0 + \sum_{k=1}^d r_k - \max_{k=1, \dots, d} (r_k) \leq \sum_{k=1}^d w_k.$$

**7.3. Non-superposition codes.** In Example 17, we presented a network in which superposition codes are non-optimal. Even though non-superposition codes are used, the tools described in Section 6 may still be used. We are going to prove the achievable region of a generalized version of Example 17.

**Example 20.** There are 3 transmitters and 3 receivers numbered 1,2,3. Receiver 1 receives from Transmitter 1 and 2, and wish to decode at  $r_1$ . Receiver 2 receives from Transmitter 1 and 3, and wish to decode at  $r_2 \geq r_1$ . Receiver 3 receives from Transmitter 2 and 3, and wish to decode at  $r_3 = r_2$ .

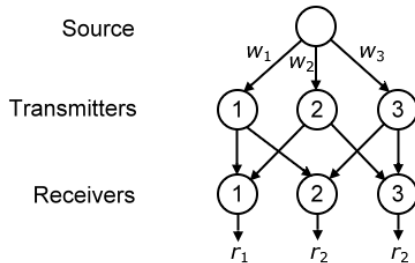


FIGURE 7.4. The network specified in Example 20

The achievable region is given by the following theorem.

**Theorem 21.** *The achievable region in Example 20 can be given by*

$$\begin{aligned} w_1 + w_2 &\geq r_1 \\ w_1 + w_3 &\geq r_2 \\ w_2 + w_3 &\geq r_2 \\ w_1 + w_2 + w_3 \cdot \frac{r_2 - r_1}{r_2} &\geq r_2. \end{aligned}$$

*Proof of converse.* The first three inequalities clearly hold. For the last inequality, consider a code in which the rate requirements are  $\epsilon$ -admissible. Receiver 1 can decode at rate  $r_1$ . By (6.3),

$$\underline{J}_0^{r_1}(Y_{(w_1, w_2, 0)}) \geq (r_1 - \epsilon)(1 - H(\epsilon)).$$

Invoking (6.1) and Lemma 14, we can obtain

$$\begin{aligned} \underline{J}_0^{r_1}(Y_{(w_1, 0, 0)}) + w_2 &\geq \underline{J}_0^{r_1}(Y_{(w_1, 0, 0)}) + \bar{J}_0^\infty(Y_{(0, w_2, 0)}) \\ &\geq \underline{J}_0^{r_1}(Y_{(w_1, w_2, 0)}) \\ (7.1) \quad &\geq (r_1 - \epsilon)(1 - H(\epsilon)). \end{aligned}$$

Receiver 2 can decode at rate  $r_2$ . By (6.3),

$$\underline{J}_{r_1}^{r_2}(Y_{(w_1, 0, w_3)}) \geq (r_2 - r_1 - \epsilon)(1 - H(\epsilon)).$$

Invoking Lemma 14, we can obtain

$$\begin{aligned} \bar{J}_{r_1}^\infty(Y_{(w_1, 0, 0)}) + \underline{J}_{r_1}^{r_2}(Y_{(0, 0, w_3)}) &\geq \underline{J}_{r_1}^{r_2}(Y_{(w_1, 0, w_3)}) \\ &\geq (r_2 - r_1 - \epsilon)(1 - H(\epsilon)). \end{aligned}$$

By Lemma 15 and (6.1),

$$w_3 \geq \underline{J}_0^{r_2}(Y_{(0, 0, w_3)}) \geq \frac{r_2}{r_2 - r_1} \cdot \underline{J}_{r_1}^{r_2}(Y_{(0, 0, w_3)}),$$

and therefore

$$(7.2) \quad \bar{J}_{r_1}^\infty(Y_{(w_1, 0, 0)}) + \frac{r_2 - r_1}{r_2} \cdot w_3 \geq (r_2 - r_1 - \epsilon)(1 - H(\epsilon)).$$

By (6.1) and (6.4),

$$\begin{aligned} w_1 &\geq \bar{J}_0^\infty(Y_{(w_1, 0, 0)}) \\ (7.3) \quad &\geq \underline{J}_0^{r_1}(Y_{(w_1, 0, 0)}) + \bar{J}_{r_1}^\infty(Y_{(w_1, 0, 0)}). \end{aligned}$$

Adding (7.1), (7.2) and (7.3),

$$w_1 + w_2 + \frac{r_2 - r_1}{r_2} \cdot w_3 \geq (r_2 - 2\epsilon)(1 - H(\epsilon)).$$

The proof can be completed by taking  $\epsilon \rightarrow 0$ .  $\square$

*Proof of achievability.* We will describe a coding scheme which can achieve the proposed region. When  $r_1 = r_2$ , we may use all the transmitters to transmit at rate  $r_1$ . When  $w_3 \geq r_2$ , then we can use Transmitter 3 alone to transmit the message at rate  $r_2$ , and Transmitter 1 together with Transmitter 2 to transmit at rate  $r_1$ . Therefore we assume  $r_1 < r_2$  and  $w_3 < r_2$ .

The code is specified by the block size  $K$ , the super-block size  $L > K$ , and the parameter  $\gamma$ . Divide the message  $\{M_i\}_{i=1,2,\dots}$  into blocks of  $K$  bits  $B_i = M_{(i-1)K+1}^{iK}$ . Further divide each block  $B_i$  into two sub-blocks  $B_{1,i}$  and  $B_{2,i}$ , where  $B_{1,i}$  contains the first  $\gamma K$  bits of the block, where  $\gamma$  is taken to be  $w_3/r_2$  in this

case, and  $B_{2,i}$  contains the rest of the bits. Group the sub-blocks into super-blocks by

$$S_{a,b,i} = (B_{b,(i-1) \cdot r_a \cdot L/K+1}, \dots, B_{b,i \cdot r_a \cdot L/K})$$

for  $a, b \in \{1, 2\}$ . Assume  $r_1 \cdot L/K$  and  $r_2 \cdot L/K$  are integers. Each super-block  $S_{a,1,i}$  contains  $r_a \cdot \gamma L$  bits, and each super-block  $S_{a,2,i}$  contains  $r_a \cdot (1 - \gamma)L$  bits.

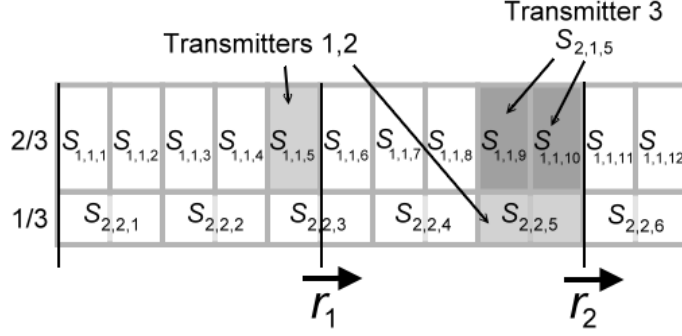


FIGURE 7.5. The super-blocks when  $\gamma = 2/3$  and  $r_2 = 2r_1$

At time  $i$ , Transmitter 1 encodes the super-blocks  $S_{1,1,\lceil i/L \rceil}$  and  $S_{2,2,\lceil i/L \rceil}$  using random linear projection (concatenate the bits in the super-blocks and transmit a random linear projection of the resultant vector). Transmitter 2 uses the same encoding scheme as Transmitter 1. Transmitter 3 encodes the super-block  $S_{2,1,\lceil i/L \rceil}$  using random linear projection.

To see why Receiver 2 can decode at rate  $r_2$ , assume that at time  $jL$ , the super-blocks  $S_{2,1,k}$  and  $S_{2,2,k}$  are already decoded for  $k = 1, \dots, j$ . During the time interval  $jL + 1, \dots, (j+1)L$ , Receiver 2 will receive  $w_3L$  bits from Transmitter 3, and  $w_1L$  bits from Transmitter 1. For a bit encoded by Transmitter 1 at time  $i$ , as the super-block  $S_{1,1,\lceil i/L \rceil}$  is already decoded (for  $j$  large enough), it can be treated as a random linear projection of  $S_{2,2,\lceil i/L \rceil}$ . Therefore all together we have  $w_3L$  projections of  $S_{2,1,j+1}$ , and  $w_1L$  projections of  $S_{2,2,j+1}$ . By definition of  $\gamma$ ,

$$w_3L = r_2\gamma L.$$

From the assumption  $w_1 + w_3 \geq r_2$ ,

$$\begin{aligned} w_1L &\geq (r_2 - w_3)L \\ &= r_2(1 - \gamma)L. \end{aligned}$$

Therefore both  $S_{2,1,j+1}$  and  $S_{2,2,j+1}$  can be decoded at time  $(j+1)K/r_2$ . By induction, Receiver 2 can decode at rate  $r_2$ . (Note that we may assume the first few blocks are decoded, as we may allocate any extra amount of time to transmit them without affecting the asymptotic behavior of the code). Similar result holds for Receiver 3.

For Receiver 1, consider the time interval  $jL+1, \dots, (j+1)L$ . In this time interval, Receiver 2 receives  $w_1L$  bits from Transmitter 1, and  $w_2L$  bits from Transmitter 2. All together there are  $(w_1 + w_2)L$  projections of  $S_{1,1,j+1}$  and  $S_{2,2,j+1}$ . The number of bits in  $S_{1,1,j+1}$  and  $S_{2,2,j+1}$  is given by

$$\begin{aligned} r_1 \cdot \gamma L + r_2 \cdot (1 - \gamma)L &= \left( \frac{r_1 w_3}{r_2} + r_2 - w_3 \right) \cdot L \\ &= \left( r_2 - w_3 \cdot \frac{r_2 - r_1}{r_2} \right) \cdot L \end{aligned}$$

which is smaller than  $(w_1 + w_2)L$  by the assumption  $w_1 + w_2 + w_3 \cdot \frac{r_2 - r_1}{r_2} \geq r_2$ . Therefore Receiver 1 is able to decode at rate  $r_1$ .  $\square$

*Remark.* Superposition codes can achieve a smaller region given by

$$\begin{aligned} w_1 + w_2 &\geq r_1 \\ w_1 + w_3 &\geq r_2 \\ w_2 + w_3 &\geq r_2 \\ w_1 + w_2 + 2w_3 \cdot \frac{r_2 - r_1}{r_2} &\geq 2r_2 - r_1. \end{aligned}$$

## 8. CONCLUSION AND DISCUSSIONS

In this report, we have investigated the achievable regions and coding schemes for multi-rate data transmission. We have shown that superposition codes are optimal for the single transmitter setting. However, in the multiple transmitter setting, there are some non-superposition codes which outperform superposition codes.

Our results can be applied in various scenarios which requires the messages to be decoded sequentially, for example, the broadcast streaming of video. The multi-rate sequential data transmission setting can also be applied on messages divided into several levels of importance. For example, in the transmission of an interlaced image file, the data corresponding to the low-resolution part is transmitted before the data corresponding to the high-resolution part. A sequential code can ensure that, even when the receiver has variable channel condition, the low-resolution part is decoded first, and therefore the receiver can display the image with lower resolution before all data are received. If the connection might be stopped at any time, using a sequential code can ensure the received message forms a continuous segment from the beginning instead of fragmented data as in Fountain codes. In the example of image transmission, if the connection is lost in the middle of the transmission, the receiver can still decode a low-resolution version of the image.

In Example 20, we have studied a particular 3-transmitter network. Further investigation on the general 3-transmitter network, and the next step,  $N$ -transmitter networks, may be carried out in the future. Ultimately, we may consider general networks of interconnected nodes instead of only two layers of nodes (transmitters and receivers). Another direction is to find a method to construct non-superposition codes according to the network connections and decoding rate requirements. Example 20 presents the construction of a non-superposition code using a sub-block structure. It is left for future studies to find out whether this construction method give the optimal code in more general settings.

## REFERENCES

- [1] R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," *IEEE Trans. Info. Theory*, pp. 609–621, 1999.
- [2] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Info. Theory*, pp. 1737–1744, 1996.